massachusetts institute of technology — artificial intelligence laboratory

Dissociated Dipoles: Image Representation via Non-local Comparisons

Benjamin J. Balas and Pawan Sinha

AI Memo 2003-018 August 2003
CBCL Memo 229

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **AUG 2003**	2. REPORT TYPE	3. DATES COVERED **00-08-2003 to 00-08-2003**

4. TITLE AND SUBTITLE
Dissociated Dipoles: Image Representation via Non-local Comparisons

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Massachusetts Institute of Technology,Artificial Intelligence Laboratory,77 Massachusetts Avenue,Cambridge,MA,02139

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**15**	

Abstract

A fundamental question in visual neuroscience is how to represent image structure. The most common representational schemes rely on differential operators that compare adjacent image regions. While well-suited to encoding local relationships, such operators have significant drawbacks. Specifically, each filter's span is confounded with the size of its sub-fields, making it difficult to compare small regions across large distances. We find that such long-distance comparisons are more tolerant to common image transformations than purely local ones, suggesting they may provide a useful vocabulary for image encoding. .

We introduce the "Dissociated Dipole," or "Sticks" operator, for encoding non-local image relationships. This operator de-couples filter span from sub-field size, enabling parametric movement between edge and region-based representation modes. We report on the perceptual plausibility of the operator, and the computational advantages of non-local encoding. Our results suggest that non-local encoding may be an effective scheme for representing image structure.

This research was funded in part by the DARPA HumanID Program and the National Science Foundation ITR Program. BJB is supported by an NIH Training Grant. PS is supported by an Alfred P. Sloan Fellowship in neuroscience and a Merck Foundation Fellowship.

The question of how to represent image content for the purposes of recognition is central to the study of human and machine vision. The challenge is to determine a vocabulary of elementary measurements that are stable across appearance variations caused by transformations such as illumination changes, defocus, translation and non-rigid deformations.[1,2] One important class of representation schemes uses aggregate statistics (such as histograms) about attributes like hue, luminance or local orientations. These schemes have proved useful in situations where object shape is likely to be highly variable, for instance, when searching for a shirt with distinctive colors in a pile of clothes[3]. However, for many settings, shape is a key determinant of object identity. Work from several laboratories[4, 5] suggests that for at least some classes of objects, surface texture and color cues provide little recognition advantage over line drawings. Consequently, much research attention has been directed towards shape-representation schemes[6].

A popular shape-representation approach involves encoding images via a collection of edge fragments along with their locations. Physiological support for this model of image representation dates back to Hubel and Wiesel's work with feline striate cortex, which revealed both "simple" and "complex" cells capable of detecting edges and lines[7]. The receptive field properties of these cells are believed to arise through a combination of aligned inputs from the LGN and intra-cortical circuitry[8-10]. Computationally, the receptive fields of these cells are commonly modeled as Gabor functions with excitatory and inhibitory lobes[11-16]. These operators may assume different orientations and positions, and also may appear at multiple scales to extract both coarse and fine structure from an image[17]**.** Such operators have been found to be useful for performing basic functions like contour localization[18]. They have also been used for more complex tasks like object detection and recognition[19-22]. Further support for Gabor wavelet operators has been provided by demonstrations that such receptive fields evolve from neural networks trained to efficiently, and with high fidelity, encode natural scenes[23].

While Gabor-like operators provide a simple means of representing image structure, the local image processing they embody limits a recognition system in some significant ways. First, edge-based representations may fail to adapt to small changes in an image brought on by changes in object geometry or position. This particular weakness stems from more general problems with edge-based algorithms, namely that most natural images contain relatively few high-frequency (edge-like) components as evidenced in the histogram of local difference values shown in figure 1. Consequently, edge maps implicitly ignore most of the content of an image, and can suffer dramatically from subtle transformations that perturb edge locations while leaving large portions of the image untouched.

Furthermore, some simple analyses also strain the capabilities of a Gabor-based representation scheme due to the conflation of the size of an operator's lobes with the distance spanned by that operator. In fact, any comparison of small regions across large distances proves quite difficult, since an operator large enough to span the relevant distance must trade resolution for size. Alternatively, comparing distant regions by propagating information via a chain of small sized operators leads to an increased susceptibility to noise contributed by each of the intermediate elements.

It is evident that the primary source of the aforementioned shortcomings of the conventional differential operators is the confounding of the inter-lobe distance with lobe-size. To overcome the shortcomings, therefore, we have to de-couple the lobe-size and inter-lobe distance parameters, thus allowing the operator to compare small regions separated by large distances.

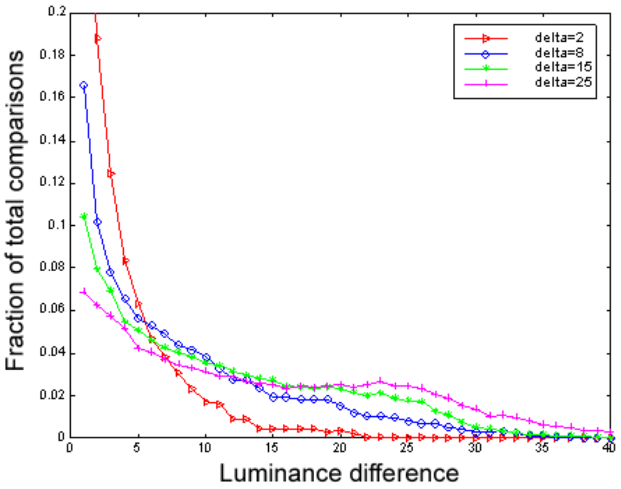

Figure 1. The distributions in a natural image of gray-level differences between pairs of pixels when only immediate neighbors are considered (delta=2) and when comparisons are made betweeen more distant neighbors only (delta=8,15,25). Due to the spatial redundancy inherent in natural images, most of the comparisons in the first case are uninformative. The resulting sparsity in responses has some advantages [Olshausen and Field, 1996], but may also lead to unstable encodings of image structure as discussed in the text.

With this motivation, we introduce the "Dissociated Dipole" or "Sticks" operator as a tool for performing non-local image comparisons. Like a simple edge-finder, a Stick is a differential operator consisting of an excitatory and an inhibitory lobe, and may be used at any orientation or scale. However, unlike a conventional edge detector, we allow an arbitrary separation between these two lobes, removing the correlation of inter-lobe distance and lobe size. Formally, the basic form of a sticks operator comprises a pair of Gaussian lobes, each with standard deviation s and a spatial separation of d. The line joining the centers of the two lobes is at angle ? relative to the horizontal, with ? ranging from 0 to 2π (figure 2). In the current implementation, we have chosen to define stick operators on image luminance values, given the significance of this attribute for form perception[24]. We note that this basic form could be altered to incorporate image attributes other than luminance, non-Gaussian lobes, or lobes of different shape and size.

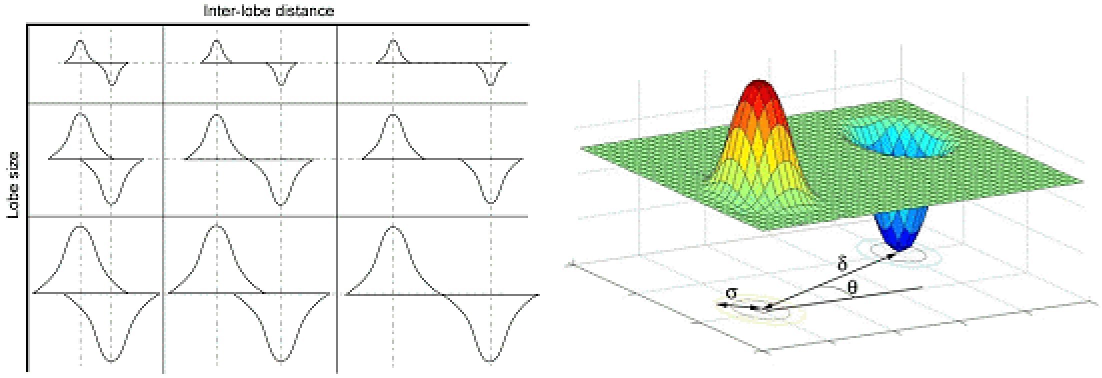

Figure 2. (Left) Conventional multi-scale representations that use Gabor-like units confound the two parameters of inter-lobe distance and lobe size. Effectively, they use only the diagonal elements of the space defined by these two parameters. The idea behind the dissociated dipoles approach is to de-couple the two attributes, in effect using the off-diagonal elements. Note that units below the diagonal have overlapping lobes and, after response cancellation, lead to adjacent lobed receptive fields. They are not considered further in this paper. We focus on the above diagonal elements which correspond to our conceptualization of dissociated dipoles. (Right) A schematic representation of a prototypical dissociated dipole.

Sticks can thus accomplish sensitive comparisons of small, distant regions while remaining agnostic about the intermediate image structure. Sticks also make more efficient use of image content than localized filters, as the probability that two distant regions in an image will differ is much greater than that of two adjacent regions, given the spatial redundancy typical of natural images[25] (figure 1,delta>2). It is important to consider why it might be useful to compare image regions separated by large distances. Preliminary support for the use of non-local measurements, besides the conventional purely local ones, comes from an examination of their tolerance to changes in image appearance. Figure 3 shows plots of the influence of a few transformations on randomly placed local and non-local differential operators. The abscissa and ordinate correspond to the response values of the operators before and after a transformation. A perfectly stable representation will lead to all points lying along a line of slope 1 passing through the origin. In general, the higher the correlation-coefficient of the scatter plot, the greater the stability. As figure 3 shows, non-local sticks-based measurements appear to be more stable against a range of common image transformations than purely local ones.

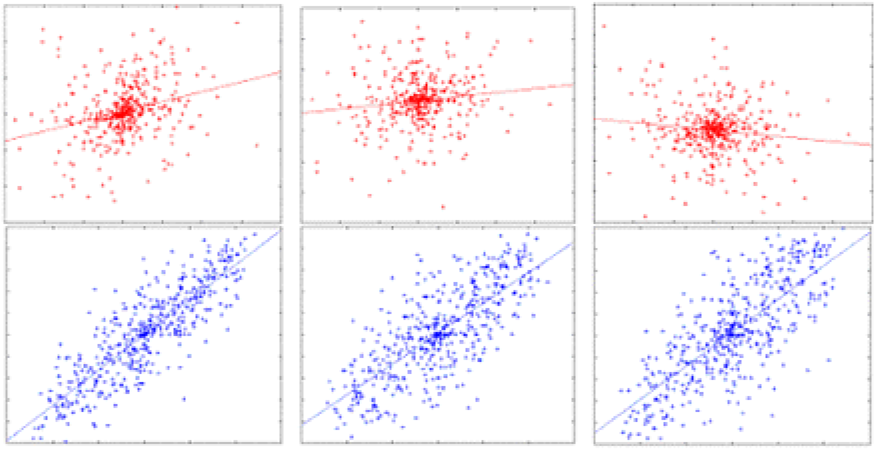

Figure 3. Comparisons of the stability of local and non-local measurements to a few image transformations. Plots of local (top) vs. non-local measurements (bottom) across changes in (left to right) expression, viewpoint and translation. The non-local measurements are more robust to these transformations as demonstrated by the values of Pearson's R obtained for each cluster of points: Expression change: local = 0.28, non-local = 0.85; Viewpoint change: local = 0.13, non-local = 0.75. Translation: local = -0.12, non-local = 0.88.

In this paper, we investigate the strengths and weaknesses of Sticks operators. We hypothesize that non-local processing (as simulated by Sticks operators) may play an important role in visual analysis. To support this hypothesis, we present results from our psychophysical studies investigating the proficiency of human observers at making non-local comparisons. Next, we demonstrate the use of the Sticks operators for complex recognition tasks through the implementation of a Sticks-based face-database search system. We examine performance of the Sticks approach across various types of image degradation and compare it with a PCA-based approach[26] to explore its relative strengths and weaknesses.

Results

Psychophysical evidence of non-local analysis

The proposal of using non-local comparisons for analyzing visual scenes is plausible only if human observers can, in fact, make such comparisons. Accordingly, we conducted psychophysical experiments designed to assess performance on this task.

Subjects performed a simple "one-up, one-down" staircase task[27,28] which required them to indicate which of two spatially separated probe regions was brighter for a range of eccentricities (0, 9.5, 16.5, 23.5 and 30 degrees of visual angle.) Each subject's chance threshold of contrast sensitivity was determined by averaging together the Michelson contrast of the stimuli from all occasions on which the subject's responses reversed from correct to incorrect (and vice versa). Four subjects participated in a version of the task that contained only horizontal separations between probes. Three of them also participated in a second task wherein the probes were separated vertically.

In both conditions, we found that subjects' ability to compare spatially distinct regions remained quite stable for large separations. For all participants in both conditions, the threshold of contrast sensitivity was located between Michelson contrasts of 0.02 and 0.08, with the steepest change in sensitivity accompanied only with the most widely separated probes. Overall, we observed that subjects were typically better at performing the task with horizontally separated probes. This difference between horizontal and vertical separation is intriguing, but not pursued further in this study. Figure 4a shows the relationship between eccentricity and contrast sensitivity for each subject, while figure 4b shows the average across all participants.

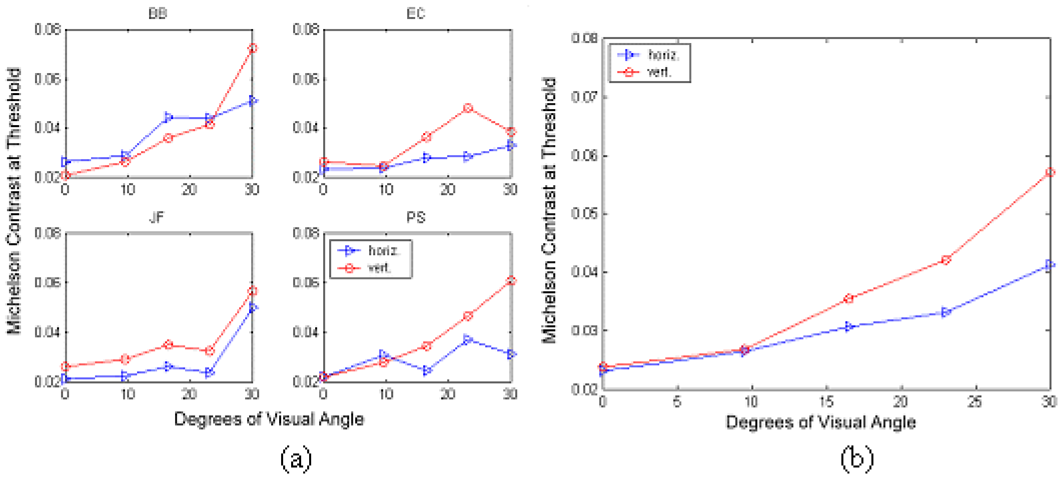

Figure 4. (a) Plots of contrast threshold v. eccentricity for all subjects in our psychophysical task. In each plot, the solid line represents performance on horizontally separated probes, while the solid line with triangles represents performance across vertical separations. (b) The average results across all subjects.

Subjects demonstrate in both tasks that they are capable of accurately performing non-local comparisons of image regions over significant separations. What might be the nature of mechanisms that underlie this ability? We consider a few possibilities. First, a large Gabor-like filter could be used to span the distance between the probes, but results from the large body of work on human contrast sensitivity suggest that subjects' ability to assess luminance differences is highly impaired beyond 0.1 cpd[29] (effectively a separation of 9.5 degrees in our setup). Given that we observe performance significantly above chance at separations beyond this limit (at least in the horizontal condition), this explanation seems inadequate. Second, sequential foveation of the probes via eye movements could render our spatial task into a temporal or even an adaptation task[30] and thereby account for the data. However, the brief presentation time and fixation requirement (see Methods) lead us to believe that this is unlikely. Finally, the results may potentially be explained by the joint operation of a chain of small operators, passing information about the probes to the middle of the chain for comparison. While theoretically possible, this scheme is susceptible to an accumulation of errors as the information propagates and also difficult to implement in the presence of any response non-linearities in the constituent units. We believe that the Sticks operator offers an explanation of these results that is both parsimonious and intuitive, compared to explanations that rely on local processing alone. Indeed, past psychophysical studies of the long-range processing of pairs of lines suggest the existence of similarly structured "coincidence detectors" which enable non-local comparisons of simple stimuli.[31-32]These detectors could make important contributions to shape representation, as demonstrated by Brubeck's idea of encoding shapes via medial "cores" constructed by integrating information across disparate "boundariness" detectors.[33]

Certainly, in the absence of direct receptive field mapping studies, the proposal of such an operator is speculative. Even if future studies reveal that the sticks operator itself does not have a physical basis, the computation it embodies - comparison of non-local regions, can still guide our attempts for devising an effective image representation strategy. In other words, measurements across distant regions may play an important role in visual analysis, irrespective of the exact mechanism by which they are extracted. We continue by examining how non-local information could contribute to complex visual tasks. We develop our ideas in the context of a content-based image retrieval system.

Non-local information benefits recognition systems

The choice of an image representation scheme greatly influences the performance of subsequent stages of visual analysis. To examine the usefulness of Sticks-based encoding to high-level vision tasks, we have implemented a system to recognize faces based on image comparisons across a range of distances (local and non-local.)

In this system, an image is represented via a "constellation" of Sticks operators. To provide a physical analogy, this is akin to dropping a number of real sticks onto an image and having them land at arbitrary orientations and positions. The difference between two small regions centered at a stick's endpoints constitutes one component of a feature vector. This feature vector then serves as the representation of a given image. To assess the similarity between two images, we simply apply the same constellation of sticks to each image, and compare the two feature vectors by a conventional distance metric, such as the L2 norm.

We used this general algorithm to assess the relevance of non-local information to recognition by conducting a series of simulations with a face database publicly available from AT&T Laboratories in Cambridge. The database contains multiple images corresponding to each of 40 individuals. The various images of an individual differ in lighting, expression and artifacts such as glasses. The simulations involved partitioning the database into mutually exclusive training and test sets. Sticks constellations were used to classify the test images. Performance was determined as a function of maximum allowable stick length.

Classification was performed in two different ways. In the first set of simulations, an L2 norm defined on the sticks feature vector was used as the measure of similarity between two images. New images were classified according to which training image they were "closest" to, and the percent of images correctly classified was obtained for any given stick constellation. In a second set of simulations, sticks feature vectors were used as the raw data for a support vector machine classifier[34] (see Methods for details.) For each of the forty individuals, the area under the ROC curve corresponding to the classification of the rest of the database was calculated. In both cases, to remove the effects of particular configurations, multiple constellations were used.

In both cases, a clear advantage for constellations of Sticks that included non-local measurements was found. A one-way ANOVA run on the SVM data showed the existence of a strong beneficial effect of including non-local information (p=0.0005) Examining figure 5a, it is apparent that the largest increases in performance were associated with an intermediate value for the upper bound of stick length. This is not surprising, as extremely long sticks can only assume a limited range of orientations and positions in an image, rendering them ineffective.

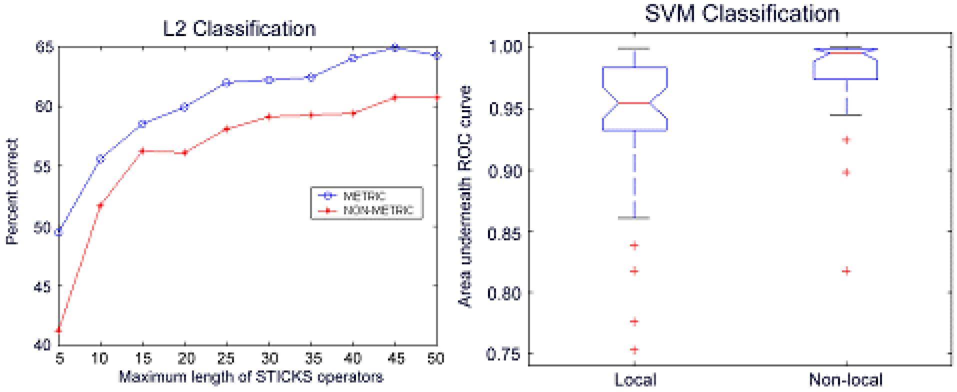

Figure 5. The results of adding non-local information to L2 (left) and SVM (right) classifications. In both cases, a clear increase in accuracy is apparent with the inclusion of non-local information. In the case of L2 classification, percent correct is used as the measure of recognition performance, while A' is used to assess our SVM simulations. The red line marked off in asterices in the L2 graph represents the performance of a "qualitative" STICKS operator that only retains the direction of polarity between two regions, instead of the precise difference magnitude.

In the interests of aligning our computational results with our psychophysical findings, we include in Figure 5a a graph representing the performance of a 'qualitative' sticks algorithm commensurate with the abilities of human observers outlined previously. In this scenario, we imagine that our observer is only capable of detecting differences in contrast above a certain threshold, and only reports the contrast polarity between two locations. Even under these circumstances, we see a clear increase in performance as maximum stick length increases. From all these results, we conclude that the inclusion of non-local image structure can substantially benefit recognition performance.

Sticks performance subject to degradation

Images in the real world are often degraded due to factors such as blur and occlusion. The ability of recognition systems to cope with noisy or incomplete images is of substantial ecological significance, and human observers are capable of performing recognition tasks despite profoundly deteriorated images[35-37]. To assess how well sticks-based encoding can handle reductions in image quality, we conducted simulations of the kind described above with inputs subjected to different degrees of degradation. Furthermore, to help contextualize the results, we compared the results with those obtained using an eigenface-based system[26]. We chose eigenfaces as a standard for comparison because it is a popular approach for face recognition, and also because it does not require the user to locate any particular features of the image prior to classification. It is important to note that the two approaches are not mutually exclusive and a hybrid system may attain even higher performance than either one alone. Both systems could be further tailored to develop and exploit models of intra- and inter-personal variance[38], but those refinements will not be considered here.

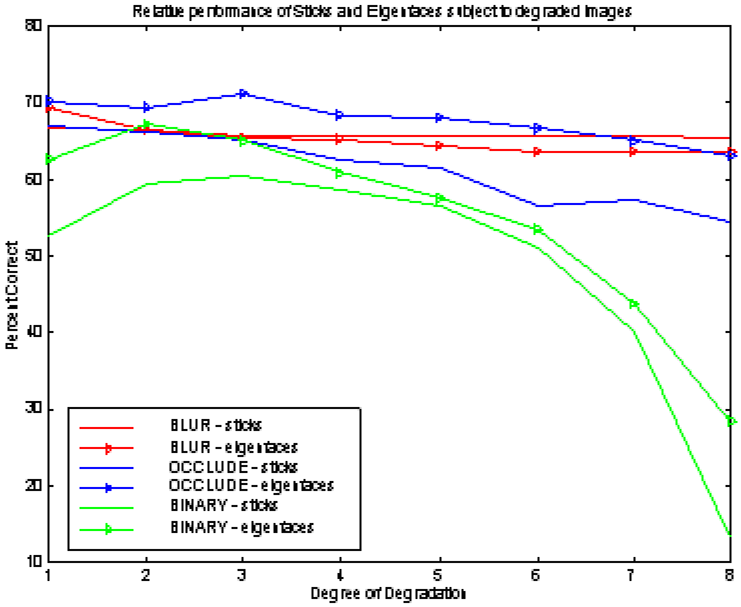

Figure 6. The performance of the sticks algorithm and the eigenfaces system subject to blurring, occlusion, and binarization.

As shown in Figure 6, the sticks approach exhibits significant robustness to degradations with very gradual fall-off of performance with increasing blur and occlusion. Performance decrement with binarization is more steep because at very high or very low thresholds, the image loses much of its structure (becoming mostly white or black). Furthermore, for most of the degraded images tested, performance of the sticks approach is comparable to that obtained with the eigenfaces system. However, with high degrees of occlusion and binarization, the latter yields better performance possibly due to a better ability to fit fragmentary image information with a holistic model. Overall, the sticks approach appears to provide robust recognition performance for a significant range of degradations. Considering the approach's simplicity and the fact that the current implementation embodies no refinements (such as better placements of sticks), we find these results very encouraging.

Discussion

Our results suggest that non-local measurements provide a useful vocabulary for encoding image structure that may facilitate recognition tasks. Moreover, our psychophysical data indicate that observers are able to make non-local comparisons in images. As for the mechanism that may underlie this ability, the sticks operator is a potential candidate.

Our proposal of such an operator does not, so far, have direct physiological support. While contextual influences on neural responses have been found in early visual areas[39-41] the form of the mapped receptive fields does not appear to correspond to a sticks operator. This is interesting given that cells with dissociated receptive field zones have been found in other sensory systems such as audition[42] and somato-sensation[43]. Some place cells in the hippocampus too exhibit multiple spatially separated 'hot spots'. This suggests that a sticks-like operator is neurally plausible.

We have presented psychophysical evidence in support of the plausibility of non-local image measurements and computational simulations that highlight their significance for robust image representation. However, these results do not allow us to comment directly on the possibility that neural systems in the mammalian visual pathway execute non-local comparisons. Previous physiological studies of early visual cortex have not yielded any cells with a receptive field structure analogous to the Sticks operator. However, we suggest that non-local processing may be accomplished by the visual system in other ways. Specifically, it has been well known for some time that long-range horizontal connections exist between cells in V1 of similar orientation tuning[44-46]. These connections bind several cells into a complex network of operators whose effective receptive field spans a region of space much larger than the constituent classical ones. This arrangement of cells in early visual cortex into a complex network may be one means by which operators in V1 combine to extract non-local relationships from an image.

The structure of this network would resemble an "extended Gabor" filter rather than a pure Sticks operator, yet changes across the network in the synaptic strength of component cells' connections could approximate variations on non-local analysis. A sub-network composed of oriented V1 cells would also permit dynamic tuning of the overall response to a gradually changing stimulus. Real-world experience is not static and unchanging, and non-local integration of cellular responses might allow for a corresponding fluidity of activation in visual cortex. Given that the functional significance of sub-networks in V1 remains unexplored, future explorations with the Sticks operator may involve examining the utility of non-local comparisons on images filtered with dynamically modifiable extended Gabor operators. More complex operators tailored to mimic the resultant filters generated by horizontal connections in V1 could also be applied to natural images as a means of determining what benefits these structures confer on recognition performance. We hope that by providing some psychophysical evidence and computational motivation for these units, this work may initiate systematic physiological investigation into the existence of such cells.

Our ongoing computational work regarding the sticks operator focuses on two questions. First, since we know the sticks algorithm does not completely encode the original image, we are currently exploring the fidelity of the sticks representation through attempts to reconstruct target images from sticks feature vectors. Previous results show that highly recognizable images can be recovered from incomplete representation schemes[47], suggesting that high fidelity may be a non-critical requirement of an image representation algorithm. Second, we are exploring whether sticks-like receptive fields emerge automatically in simulations such as those used by Olshausen and Field[23], when the criterion of performance is not the quality of reconstruction but stability to transformations. Preliminary results (Figure 7) suggest that a range of operators, local and non-local, emerge naturally given these constraints. These results also indicate that it might be imprudent to set up a forced dichotomy between representations based exclusively on local operators and those based on non-local ones. A scheme that combines both kinds of analyses might yield better results than either one alone.

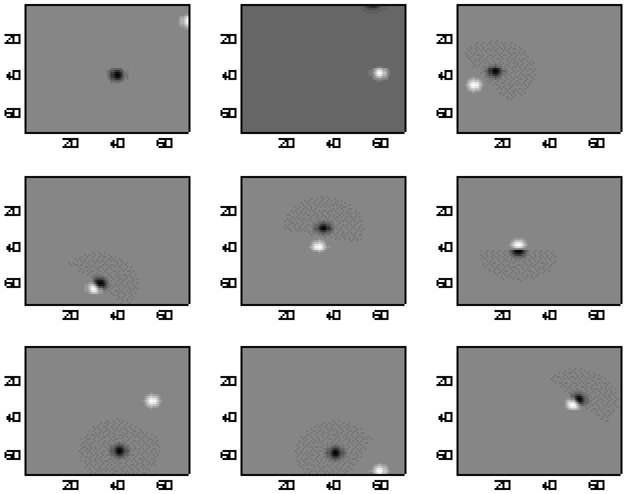

Figure 7. A family of receptive fields found by optimizing the locations of two Gaussians for the purposes of distinguishing between two individuals. Both local and non-local operators are evident in this series of plots.

Methods

Psychophysics

The authors (BB and PS) and two naïve subjects participated in this task. Subjects had normal or corrected to normal acuity. In all experiments, we used a chin mount to stabilize subjects' heads and facilitate consistent central fixation. Subjects viewed the stimuli monocularly with the open eye centered with respect to the screen (where the fixation mark appeared), and responded to each trial via keyboard presses. Experimental sessions were conducted individually for each subject.

Our stimuli comprised pairs of simultaneously presented small image patches (probes), wherein each probe could be assigned an arbitrary luminance. Each probe was a square subtending approximately 2 degrees of visual angle. Subjects' task was to determine which of the probes was lighter on any given trial while maintaining fixation on a centrally presented small red dot. The probes were placed symmetrically relative to the fixation mark and the distance between the probes could be 0, 9.5, 16.5, 23.5, or 30 degrees of visual angle on any given trial. We experimented with both horizontal and vertical placements of the probes. The stimulus eccentricity used on a given trial was selected randomly from the set of five possibilities. This was to discourage subjects from making predictive eye-movements or allocating attention to specific locations. The position of the lighter probe (left/right, top/bottom) also varied randomly. Stimulus presentation time was limited to 200 ms to further ensure that subjects were not able to make alternating eye movements to the two probes perform the task. Since the display during the inter-stimulus period comprised just the fixation dot on a blank field, we were concerned that subjects may begin losing sensitivity to peripheral field due to image stabilization. To address this problem, we followed each stimulus presentation by two wide-field noise masks, each displayed for 200 ms. A new trial was initiated after the subject had responded to the previous one. Subjects were shown 40 trials for each separation magnitude, for a grand total of 200 trials per subject in each of the horizontal and vertical conditions.

We employed a simple staircase procedure (one-up, one-down) to locate the contrast level at which subjects were performing at the level of chance (50% for this 2AFC task). With each correct answer, the intensity of the dark probe was increased by 0.5 cd/m^2 and the light probe's intensity was decreased by the same amount. Incorrect answers led to a change in probe intensity of the same magnitude, but opposite sign (thus increasing the luminance difference). The light and dark probes were initially presented to the subject with intensities of 46 cd/m^2 and 26 cd/m^2 respectively, against a white (120 cd/m^2) background. At the lowest possible level of contrast, the probe intensities were 37 and 36 cd/m^2. The task was self-paced, although subjects were encouraged to respond as quickly and accurately as possible. All four subjects participated in the horizontal probe condition and three of them (one male, two females) also participated in the vertical probe condition.

Image Set

The images used throughout this paper are taken from the ORL face database, which contains ten grayscale images each of forty individuals. The images are 112x92 pixels in size, and vary in face position, expression, and some amount of depth rotation. All images were preprocessed to have zero mean and unit variance.

The Basic Sticks Algorithm

Sticks are "dropped" onto an image by first choosing one image point at random for each operator. A second point for each operator is located by moving +/- X units away from the first point horizontally, and +/- Y units away vertically. X and Y are each taken from independent and uniform distributions between 0 and an upper bound specified by the user. Unless otherwise specified, the upper bound on X and Y in all of our simulations was 50 pixels, yielding a range of stick lengths between 0 and ~70 pixels.

A square "lobe" is built around each point of a sticks operator by averaging the intensity values within a window of R units above, below, left and right of each sticks point, where R may be specified by the user. (Unless otherwise specified, we use a value of 2 pixels for R in all cases.) The difference between the value of the first lobe and the second lobe is the operator's output. "Noisy" perception may be modeled by incorporating a difference threshold that limits the ability of the sticks algorithm to perceive contrast between its end points. If the absolute difference between lobes exceeds this threshold, the value of that stick is either 1 or –1, depending on the direction of contrast. Otherwise, the value of that operator is 0. We used a difference threshold of 30 gray levels (0-255 scale) for our "non-metric" simulations.

Finally, a feature vector is created by combining the output of all sticks operators in a constellation into one list. This feature vector is the final output of the sticks algorithm for a particular image, and the image encoding that we use for classification.

Classification Methods

We chose to use L2 norms and SVM tools to classify sticks feature vectors obtained from the ORL database. For all L2 simulations, the first image of each individual in the ORL database was used for training, with one constellation of 150 sticks used to extract feature vectors for each image. Subsequent images were transformed into feature vectors via the same constellation, and classified according to the minimum Euclidean distance to one of the training images. Classification was deemed successful if the training image selected depicted the same individual as the input.

SVM classification was carried out with the LS-SVM package. In this case, only 50 sticks were used to create each feature vector, and the first 5 images of each individual were used as training data. The remainder of the database was used as a test set, where the task was to classify all new images as either new views of one particular individual or of another person. For each constellation, this task was performed for each of the 40 individuals in the database, and the area under the ROC curve was used as a measure of performance. The average performance per individual achieved across 20 different sticks constellations was used to wash out the effects of particular constellations, and an RBF kernel (gamma=10, sigma=3) was chosen for classification.

Finally, to classify images via eigenfaces, the images used as a training set for L2 sticks classification served to provide a basis for Principal Components Analysis via Matlab's SVD function. Each training image was then projected onto that basis, with new images classified according to the minimum L2 distance to a training image in the new "face space."

Image Degradations

We carried out blurring by convolving a Gaussian filter of increasing variance with the input image. Increasing numbers of 3x3 pixel medium-gray squares were randomly placed on the image to approximate scattered occluders, and binarization was achieved by thresholding the image at varying gray-levels.

Emergent Operators for Recognition

To determine what operators are best for recognition purposes, a set of 20 images from the ORL database (2 individuals, 10 images each) were cropped to exclude external features, and used as the basis for exploring the properties of two-lobed RFs. Two Gaussians (st. dev. = 2) were placed at the same point in an image, and were allowed to independently roam away from that starting point to locations that maximized the variance of their difference between images of different individuals, but not within different images of the same individual. ANOVA was used to determine the strength of each configuration of the two lobes, with low p-values for the effect of "different individual" being sought after as long as the p-values associated with "different image" were greater than 0.4.

References

1. Papathomas, T. V. (1995). *Early Vision and Beyond*, MIT Press, Cambridge.

2. Ullman, S., Vidal-Nanquet, M. & Sali, E. Visual features of intermediate complexity and their use in classification. *Nature Neuroscience,* **5**, 682 - 687 (2002)

3. Swain, M.J., & Ballard, D.H. (1991) Color Indexing. *International Journal of Computer Vision,* **7(1)**, 11-32.

4. Biederman, I., & Ju, G. (1988). Surface vs. Edge-Based Determinants of Visual Recognition. *Cognitive Psychology*, **20**, 38-64.

5. Delorme, A., Richard, G., and Fabre-Thorpe, M. (2000). Ultra-rapid categorization of natural scenes does not rely on color cues: a study in monkeys and humans. *Vision Research,* **40**, 2187-2200.

6. Edelman, S. (1999). *Representation and recognition in vision*. MIT Press, Cambridge.

7. Hubel, D. & Wiesel, T. (1959). Receptive Fields of Single Neurons in the Cat's Striate Cortex. *J. Physiol.***148**, 574-591.

8. Ferster, D. Chung, S. and Wheat, H. (1996) Orientation selectivity of synaptic input from the lateral geniculate nucleus to simple cells of the cat visual cortex. *Nature* **380**, 249-252.

9. Somers, David C., Nelson, Sacha B., and Sur, Mriganka (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *Journal of Neuroscience*, **15**, 5448-5465.

10. Troyer, T.W., A.E. Krukowski and K.D. Miller (2002). *LGN input to simple cells and contrast-invariant orientation tuning: An analysis. J Neurophysiol***.** **87**, 2741-2752.

11. Koenderink, J. J. & van Doorn, A. J. (1976). Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics*, **21(1)**, 29-35.

12. Malik, J. & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America,* **7(5)**, 923-932.

13. Daugman, J. G. (1990). An information theoretic view of analog representation in striate cortex. *Computational Neuroscience,* MIT Press, 403-424.

14. Daugman, J. G. (1997). Complete discrete 2D Gabor transforms by neural networks for image analysis and compression. IEEE Transactions on Acoustics, Speech and Signal Processing, **36(7)**, 1169-1179.

15. Jones, D. G. and Malik, J. (1992). Computational framework for determining stereo-correspondence from a set of linear spatial filters. Image Vision Computation, **10**, 699-708.

16. Field, D. J. (1994). What is the goal of sensory coding? Neural Computation, **6**, 559-601.

17. J. Touryan and Y. Dan (2001). Analysis of sensory coding with complex stimuli. *Curr. Opin. Neurobiol.* **11**, 443-448

18. Rivest, J., & Cavanagh, P. (1995). Localizing contours defined by more than one attribute. *Vision Research* , **36**, 53-66.

19. Rao, R. P. N. & Ballard, D. H. (1995). An active vision architecture based on iconic image representations, Artificial Intelligence, 461-505.

20. Wiskott, L., Fellous, J. M., Kruger, N. & von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching, IEEE Transactions on Pattern Analysis and Machine Intelligence, **19(7)**, 775-779.

21. Shams, L & von der Malsburg, C. (2002) The Role of Complex Cells in Object Recognition. *Vision Research,* **42**, 2547-2554.

22. Viola, P. & Jones, M. (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. *Accepted Conference on Computer Vision and Pattern Recognition 2001.*

23. Olshausen, B.A. (1996). Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images. *Nature,* **381**, 607-609.

24. Shioiri, S., & Cavanagh, P. (1992). Achromatic form perception is based on luminance not brightness. *Journal of the Optical Society of America A* **9** , 1672-1681.

25. Kersten, D. Predictability and redundancy of natural images. (1987) *Journal of the Optical Society of America A,* **4**, 2395-2400. Reprinted in: Image Compression, Rabbani, M. (Ed.), SPIE-The International Society for Optical Engineering. 1992.

26. Turk, M.A. & Pentland, A.P. (1991). Eigenfaces for Recognition. *Journal of Cognitive Neuroscience,* **3(1)**, 71-86.

27. Cornsweet, T. (1962) *American Journal of Psychology,* **75**, 485-491.

28. Treutwein, B. (1995) Adaptive Psychophysical Procedures. *Vision Research,* **35(7)**, 2503-2522.

29. Owsley, C. Sekuler, R. & Siemensen, D. (1983) Contrast Sensitivity Throughout Adulthood. *Vision Research,* **23(7)**, 689-699.

30. Shapley, R., and C. Enroth-Cugell. 1984. Visual adaptation and retinal gain controls. In: *Progress in Retinal Research,* 3:263-346. Elmsford, New York: Pergamon Books, Inc.

31. Morgan, M. J., & Regan, D. (1987) Opponent model for line interval discrimination: Interval and vernier performance compared. *Vision Research,* **27(1)**, 107-118.

32. Kohly, R.P. & Regan, D. (2000) Coincidence detectors: Visual processing of a pair of lines and implications for shape discrimination. *Vision Research*, **40(17)**, 59-74.

33. Burbeck, C.A., & Pizer, S.M. (1995) Object representation by cores: Identifying and representing primitive spatial regions. *Vision Research,* **35(13)**, 1917-1930.

34. Heisele, B., P. Ho and T. Poggio. Face Recognition with Support Vector Machines: Global Versus Component-based Approach, *International Conference on Computer Vision (ICCV'01),* Vancouver, Canada, Vol. 2, 688-694, 2001.

35. Harmon, L. D. and Julesz, B. (1973). Masking in visual recognition: Effects of two-dimensional filtered noise. *Science*, **180(4091)**, 1194-1196.

36. Yip, A. and Sinha, P. (2002). Contribution of color to face recognition. *Perception,* Vol. **31**, 995-1003.

37. Sinha, P. (2002). Recognizing complex patterns. *Nature Neuroscience***, Vol. 5 (suppl.)**, 1093-1097.

38. Moghaddam, B., Jebara, T. & Pentland, A. (2000). Bayesian Face Recognition. *Pattern Recognition,* **33(11),**1771-1782.

39. Zipser, K., Lamme, V. A. & Schiller, P. H. (1996). Contextual modulation in the primary visual cortex. *Journal of Neuroscience***, 16(22)**, 7376-7389.

40. Ito M. & Gilbert, C.D. (1999). Attention modulates contextual influences in the primary visual cortex of alert monkeys. Neuron, **22**, 593-604.

41. Dragoi, V. & Sur, M. (2000). Dynamic properties of recurrent inhibition in primary visual cortex: Contrast and orientation dependence of contextual effects. J. Neurophysiol. **83**, 1019-1030.

42. Young, E.D. (1984). Response Characteristics of Neurons of the Cochlear Nucleus. *Hearing Science Recent Advances,* pp. 423-60. (C.I. Berlin, ed.) College Hill Press, San Diego.

43. Chapin, J.K. (1986). Laminar Differences in Sizes, Shapes, and Response Profiles of Cutaneous Receptive Fields in the Rat SI Cortex. *Experimental Brain Research***, 62**, 549-559.

44. Das, A. and Gilbert, C.D. (1995). Long-range horizontal connections and their role in cortical reorganization revealed by optical imaging of cat primary visual cortex. Nature**, 375**, 780-784.

45. Trachtenberg, J.T. and Stryker, M.P. (2001) Rapid anatomical plasticity of horizontal connections in developing visual cortex. **J. Neurosci**. **21**, 3476-3482.

46. Chisum, HJ, Mooser, F, and Fitzpatrick,D (2003). Emergent Properties of Layer 2/3 Neurons Reflect the Collinear Arrangement of Horizontal Connections in Tree Shrew Visual Cortex, J Neurosci. **23**, 2947-2960.

47. Sadr, J., Mukherjee, S., Toreaz, K., & Sinha, P. (2002) The fidelity of local ordinal encoding. *Advances in Neural Information Processing 14,* MIT Press.